Design and Applications of an Intelligent Financial Reporting and Auditing Agent with Net Knowledge (FRAANK)

Alexander Kogan <kogan@rutcor.rutgers.edu>
Department of Accounting and Information Systems
Faculty of Management, Rutgers University
180 University Avenue, Newark, NJ 07102-1895

Kay Nelson <nelson_k@cob.osu.edu>
Department of Accounting and MIS
Fisher College of Business, The Ohio State University
2100 Neil Avenue, Columbus, OH 43210

Rajendra P. Srivastava <rsrivastava@ukans.edu>
Ernst & Young Center for Auditing Research and Advanced Technology
Division of Accounting and Information Systems
School of Business, The University of Kansas
Lawrence, KS 66045

Miklos A. Vasarhelyi <miklosv@andromeda.rutgers.edu>
Department of Accounting and Information Systems
Faculty of Management, Rutgers University
180 University Avenue, Newark, NJ 07102-1895

Matthew Bovee <mbovee@ukans.edu>
Graduate Student, Division of Accounting and Information Systems
School of Business, The University of Kansas
Lawrence, KS 66045

Design and Applications of an Intelligent Financial Reporting and Auditing Agent with Net Knowledge (FRAANK)

Abstract

This paper discusses the use of intelligent Internet agents as essential tools for automating financial analysis functions in the virtual world. Several important characteristic features of intelligent agents are discussed, including autonomy, communication ability, collaboration, and mobility. This paper focuses on developing a new intelligent agent called FRAANK – Financial Reporting and Auditing Agent with Net Knowledge. The prototype of FRAANK described in this paper provides intelligent access to, and processing and integration of rapidly changing financial information available from various sources on the Internet. FRAANK is an example of an agent that provides a value-added service that can be used for extracting data from natural text financial statements and converting them into XBRL-tagged statements, can potentially be utilized in an auditing practice, or used by investors and creditors in making their decisions.

1 Introduction

The accounting world has dramatically changed over the recent years and the pace of change is accelerating. Major accounting firms have morphed into "professional service organizations", which have increasingly used technology to obtain labor savings and are progressively developing large knowledge management services to support their professional activities. These knowledge management activities typically draw from a mix of acquired databases (e.g., COMPUSTAT, IBES), statute information (e.g., FASB / AUDSAC), and evidential documentation from professional work that provides a base of actual firm-wide experiences. Professionals also have access to the Internet, which has exponentially increased the speed and availability of current information.

The American Institute of Certified Public Accountants (AICPA) Special Committee on Assurance Services developed a strategic plan for identifying and implementing an expanded assurance function resulting from the information technology revolution (AICPA SCAS [3], see also [13,30]). The committee has proposed the development of a series of expanded assurance services, which are defined in the following generic way (see Elliot, 1997, p. 63 [13]): "Assurance services are independent professional services that improve the quality of information, or its context, for decision makers." Several of these new assurance services, are currently being developed by committees and task forces of the AIPCA and CICA. CPA WebTrust (AICPA WebTrust [4], Greenstein [25]) and SysTrust [7] were the first members of this family.

The SYSTRUST committee emphasized the role of intelligent agent technology in supporting these new assurance services. In particular, the description of systems reliability assurance states:

> "Assurers will use audit software agents to search for unusual patterns and/or corroborative patterns in transactions. … Agents may have adaptive, quasi-learning algorithms embedded to adjust to a constantly changing model."[1]

This paper focuses on developing a new intelligent agent called FRAANK – Financial Reporting and Auditing Agent with Net Knowledge. The prototype of FRAANK described in this paper provides intelligent access to, and processing and integration of rapidly changing financial information available from various sources on the Internet. FRAANK is an example of an agent that provides a value-added service that can be used for extracting data from natural text financial statements and converting them into XBRL[2]-tagged statements, can potentially be utilized in an auditing practice, or used by investors and creditors in making their decisions. It can help improve the quality and accuracy of professional firm knowledge systems, and can also be viewed as an enabling technology for the broadly viewed assurance services discussed above. Additionally, this intelligent agent technology may improve the economics of traditional audit and serve as an integral part of the emerging area of continuous auditing pioneered by Vasarhelyi and Halper [49].

In the next section we discuss the general issues of intelligent agency in the virtual world environment. Then, in Section 3 we define the goals of FRAANK, and describe its function and architecture. In Section 4 we give some technical details about the implementation of FRAANK, while in Section 5 we discuss the relationship of

[1] http://www.aicpa.org/assurance/scas/newsvs/reliab/index.htm
[2] XBRL (Extensible Business Reporting Language) is a recent XML derivative language aimed at enhancing and facilitating the business reporting and analysis process at the account level. Several other

FRAANK with XBRL. We devote Section 6 to the description of various applications of FRAANK. In Section 7 we outline most important future developments. Finally, we provide concluding remarks in Section 8.

2 Intelligent Agents

The Virtual World Environment

The tremendous success of the Internet has caused most companies to establish online presence and continuously increase the amount of information provided online. Moreover, many companies are now redesigning their business processes around the Internet. This development will result in the possibility to provide authorized parties with online access to most corporate data. At the same time, numerous valuable data sources are already publicly available on the Internet. The most notable one is the Securities and Exchange Commission's EDGAR (Electronic Data Gathering, Analysis, and Retrieval [11]) Internet repository containing corporate filings with the SEC. Many Internet portals (e.g., Yahoo [56], Quicken [42]) provide public access to financial markets data, analysts' forecasts, news feeds of business relevance, etc. This rich virtual world environment already provides a fertile ground for intelligent software agents (see e.g., [9,10,16,17,18,33,34,37]).

Essential Features of Intelligent Software Agents

While software agents have been the focus of much attention both in popular

---

XML derivative languages exist, proposed in different industries or by associations focus on the transaction level. See http://www.xbrl.org.

press and in the research community, there is no commonly accepted definition of an agent. For an extensive discussion of what an agent is, see Franklin et al. [23]. The most essential feature of an agent is its *autonomy*, i.e., the ability to function and accomplish tasks without direct human intervention. Although autonomy may be considered as the defining feature of an agent, this feature by itself would qualify as an agent such an unsophisticated program as a remover of temporary files, which is automatically run daily. A truly autonomous agent should also be both *reactive* and *proactive*. The reactivity of an agent encompasses sensing the environment and reacting to changes in it, while a proactive agent will exhibit functions not limited to reactions but aimed at achieving its own objectives.

The complexity of tasks facing future accounting and auditing agents will require a sophisticated software system that can accomplish such task utilizing specialized knowledge about the universe it operates in. We therefore focus our attention on *intelligent* agents [54,55], since intelligence will be crucial for an agent to be able to accomplish non-trivial objectives. Only intelligent agents can be useful enough in real world scenarios. The virtual environment in which the agent will have to operate is very complex, uncertain and rapidly changing. An autonomous system without a significant level of intelligence will not be able to function successfully in such an environment [41]. It has to be remarked that after more than half a century of continuous research efforts in the field of artificial intelligence, no definition of what intelligence of a software system is has been universally accepted despite numerous attempts by some of the best minds in the field (e.g., the famous Turing test). It is therefore not surprising that the identification of the most important intelligent features for an accounting and auditing agent presents a challenging research problem. Past experience with various artificial intelligence applications has shown that to be successful in practice an application usually has to be sufficiently focused on a limited set of tasks. The agent we describe in this paper is focused on the task of finding important information (e.g., accounting numbers) in semi-structured natural text documents of accounting or auditing nature. The pioneering research in developing artificial intelligence capabilities for analyzing financial statements and understanding accounting texts [47] is described by Mui and McCarthy [35] and O'Leary et al. [38,39,40]. The most comprehensive implementation of the intelligent parsing of financial statements in the SEC EDGAR filings has been achieved in EdgarScan [12,21,45,46].

Communication, Collaboration and Mobility of Agents

Intelligent software agents will arguably need certain additional capabilities for many auditing and financial reporting applications. In particular, such agents may have to be able to interact with other agents [26]. This *communication* ability may or may not be required depending on the nature of the agent's tasks. A communicating agent can be further made capable of taking part in *collaboration* with other agents. Such collaboration can only be possible if all the participating agents adhere to the same standards in their communications, or if a specialized intermediary agent provides inter-agent translation. If the FRAANK agent described in this paper is furnished with communication capabilities, then it can field requests from and provide parsed financial data back to other agents, e.g., those agents that automate various applications described in Section 6, such as financial analysis, credit worthiness evaluation, or analytical review procedures using peer companies financial data.

The initial attempts to standardize an agent communication language (ACL) began as a part of the ARPA Knowledge Sharing Effort, which resulted in the development of the Knowledge Query and Manipulation Language (KQML) [48]. According to Singh [43], the success of KQML has been limited, and a new competing standard for ACL is being developed by the Foundation for Intelligent Physical Agents (FIPA) [22].

The standard for the representation of information exchanged in inter-agent communications is the Knowledge Interchange Format (KIF); see Genesereth and Ketchpel [24]. Collaborating agents can form multiagent systems, which can be essential for coordinating and integrating processes in a virtual enterprise [28].

As the last but not least important feature of an agent, we mention *mobility*. In certain cases an accounting and auditing agent may need mobility for transporting itself or its offspring to remote locations and executing some tasks there. The Internet provides a natural environment for mobile agents [27,29,32,53], as an ever-growing number of organizations recreate themselves in the virtual world. The quest for mobility emerges when the amount of information required by the agent is so vast that it would be either technologically infeasible or prohibitively expensive to transfer that information over the network. In this case, instead of moving the mountain of information, the agent will move itself to that mountain, and process information there. Additionally, mobility becomes essential whenever the latency of the network connection severely impacts the agent's ability to generate real-time responses, or whenever the network connectivity is sporadic.

We view intelligent agents as essential tools for future accounting and auditing applications. Intelligent agents possess the potential to both automate many existing routines and at the same time enrich the current practice by providing new capabilities that could not have been implemented otherwise.

3 Goal, Function and Architecture of FRAANK

The ultimate goal of the Financial Reporting and Auditing Agent with Net Knowledge project is the creation of a universal financial reporting and audit assistant capable of:

- interacting with a variety of information sources;
- learning the structure of on-line sites;
- analyzing natural text in the accounting and audit domains; and
- formally representing and using accounting and audit knowledge.

The FRAANK prototype can be viewed as the first step in gathering financial data on companies that is available on the Internet and using this data to provide value added service. Once the agent brings financial information back to the home computer, this data can be processed and combined with other artificial intelligence systems to provide knowledge for enhanced decision-making in real time mode. The prototype of the FRAANK agent can be found at http://lark.cc.ukans.edu/cgiwrap/srivasta/agent.cgi or http://fraank.eycarat.ukans.edu/cgi-bin/agent.pl

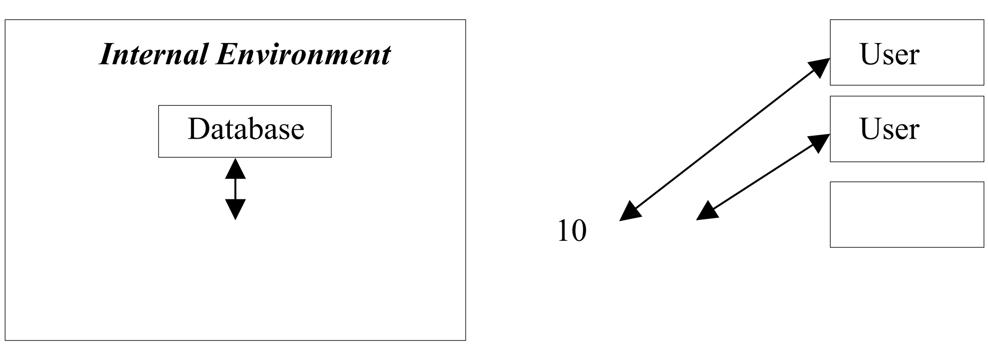
FRAANK communicates with its users over the World Wide Web through a simple interface. The user, who is interested in a certain publicly traded company, submits that company's name to the agent, and then FRAANK automatically:

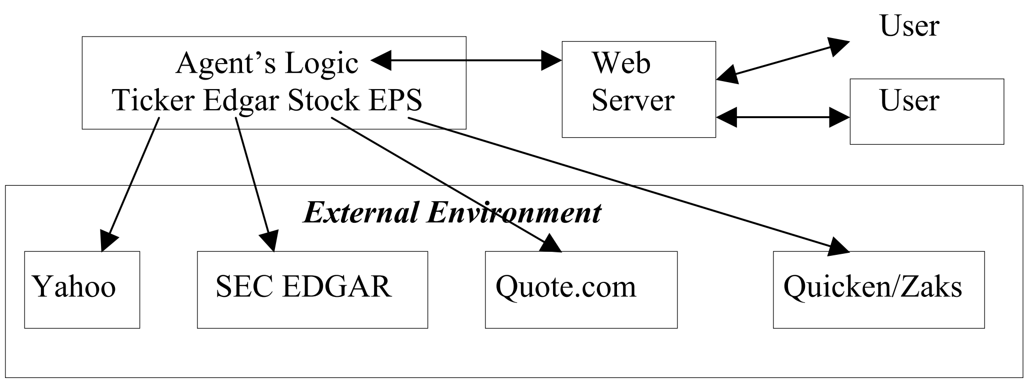
- searches the SEC EDGAR database for the quarterly filings (10-Q) by that company,
- queries the Yahoo ticker search engine to identify the ticker of the company,
- uses the ticker to retrieve the most recent stock price from Quote.com,
- contacts Quicken to find out the most current consensus forecast of earnings per share (EPS) provided by Zacks, and
- utilizes the obtained results to calculate various accounting ratios, the Z-factor (a discriminant measure of bankruptcy risk) (Altman [1,2]), and lambda (λ, a stochastic measure of solvency)(Emery & Cogger [14]).

Architecture of FRAANK

The multi-faceted nature of FRAANK is reflected in its design. FRAANK consists of several sub-agents, corresponding to the Internet information sources that FRAANK utilizes in real-time. These sub-agents are the Edgar agent, the ticker agent, the stock quote agent, and the EPS agent. The architecture of FRAANK, as shown in Figure 1, allows the agent's logic to be clearly separated from the end user interface on the one hand, and from the accounting knowledge source stored in a relational database on the other hand. Currently, the knowledge source in the database contains synonyms of accounting terms.

Figure 1: The Architecture of FRAANK





We will discuss next the functionality of the FRAANK sub-agents.

Edgar

Edgar is the main sub-agent of FRAANK, since it is responsible for obtaining and processing the richest source of information – quarterly filings of publicly traded companies. The Edgar sub-agent is described in detail in K.M. Nelson et al. [36]. The main functions of Edgar include:

- retrieving the company's recent quarterly SEC filings,
- finding and extracting from them consolidated financial statements, and
- parsing the statements for important accounting numbers.

This sub-agent is called the Edgar agent since it searches the SEC EDGAR database for quarterly reports (10-Q) of publicly traded companies for matches to the name of a user-specified corporation and returns a list of available 10-Q filings. These filings are sorted by date with the most recent ones listed first. The user is prompted to choose a 10-Q filing, which is then retrieved and analyzed. Since the Edgar agent always contacts the SEC EDGAR database in real-time over the Internet, it always provides access to the most recent filings. While this retrieval function is based on the standard Internet technology, the other two functions of Edgar have to do with extremely difficult and highly non-standard problems of analyzing natural text.

The current version of FRAANK finds in the retrieved 10-Q filing the consolidated balance sheet, income statement, and statement of cash flows. Although a human accountant can easily locate a financial statement in the body of a quarterly filing, the same task is a challenge for a computer program because of the great variability in the structure of the filings and the language used. The location of the statements in the body of the filing varies significantly. While each statement is delimited with the table tags, there are usually many different tables in the body of the filing. Moreover, the caption can be both before and after the table tag, and the wording of the caption is unpredictable. For example, the income statement can also be called the statement of income, the statement of changes in financial position, the profit and loss statement, or the statement of revenues and expenses, etc. Complicating the automatic location of statements even further is the possibility of having the same keywords in the caption of some other tables in the body of the filing (e.g., due to the segment reporting requirements).

After a particular financial statement is extracted, it is parsed by the Edgar agent for certain line items that contain important accounting numbers such cash, total assets, total revenues, net income, etc. Line item names can exhibit even greater variability than the table captions, as described in detail in [36]. The intelligent parsing procedure in

Edgar utilizes tables of line item synonyms[3] stored in a relational database (see Section 4 for a more detailed description). It used to be possible to cross-validate certain line items with the SGML-tagged values presented in the Financial Data Schedule (FDS) appended to the 10-Q filing. Unfortunately, for undisclosed reasons, the SEC abandoned the requirement to append the FDS as of early 2001, thus making this cross-validation impossible for newer filings. Therefore, for recent filings, only the aggregation structure (totals, subtotals, etc.) of the line numbers is used to improve the accuracy of parsing.

Ticker

After the user selects a 10-Q filing for analyses, FRAANK uses its ticker sub-agent to identify the ticker symbol of the company, which is needed for obtaining the stock quotes and EPS forecasts. There are a number of ticker search engines publicly available on the Internet. The ticker agent sends the company name to the Yahoo ticker engine [56], and retrieves a list of tickers matching the submitted name. The retrieved ticker list must then be analyzed to determine the most probable candidate, since this list will usually contain quite a few tickers. The ticker sub-agent utilizes several heuristic rules for identifying the best match, applying them in the following order:

- Since most probable matches should include user input as separate words in company names, items not containing all the exact words (e.g., containing super-words) are eliminated.
- Since most probable ticker symbols are short, the list is ordered by ticker symbol length.

[3] This same method can be adapted to the creation of taxonomies for XBRL statements.

- Since most probable company names are short, items corresponding to the shortest tickers are ordered by the total length of the company name, and the first one is chosen as the best match (any tie is broken arbitrarily at this stage).

Although these rules are quite simple, and thus the ticker agent intelligence rudimentary, very often the resulting best match is exactly the right ticker.

Stock Quotes and EPS forecasts

Using the ticker symbol identified by the ticker sub-agent, the stock quote sub-agent sends the ticker to the public stock quote engines (mach.quote.com), and extracts from the response the most recent share price. Intuit on its Quicken Web site [42] provides free access to Zacks Investment Research data of earnings per share (EPS) forecasts. The EPS sub-agent uses the most probable ticker to retrieve the (diluted) EPS forecast for the next quarter from quicken.com.

Integration and Analysis

FRAANK integrates the sub-agents' responses (price from the stock quote sub-agent, earnings from Edgar, and analyst forecasts) to compute historical and estimated important financial ratios showing this company's investment potential (like the price/earnings ratio), and to calculate a number of other financial ratios (e.g. quick ratio, current ratio). As examples of summarizing and integrating the information obtained by the sub-agents, FRAANK computes two measures of the company's financial health: Altman's Z-factor (Altman [1,2]) and lambda (λ) (Emery & Cogger [14]), detailed further in Section 6.

4 Implementation of FRAANK

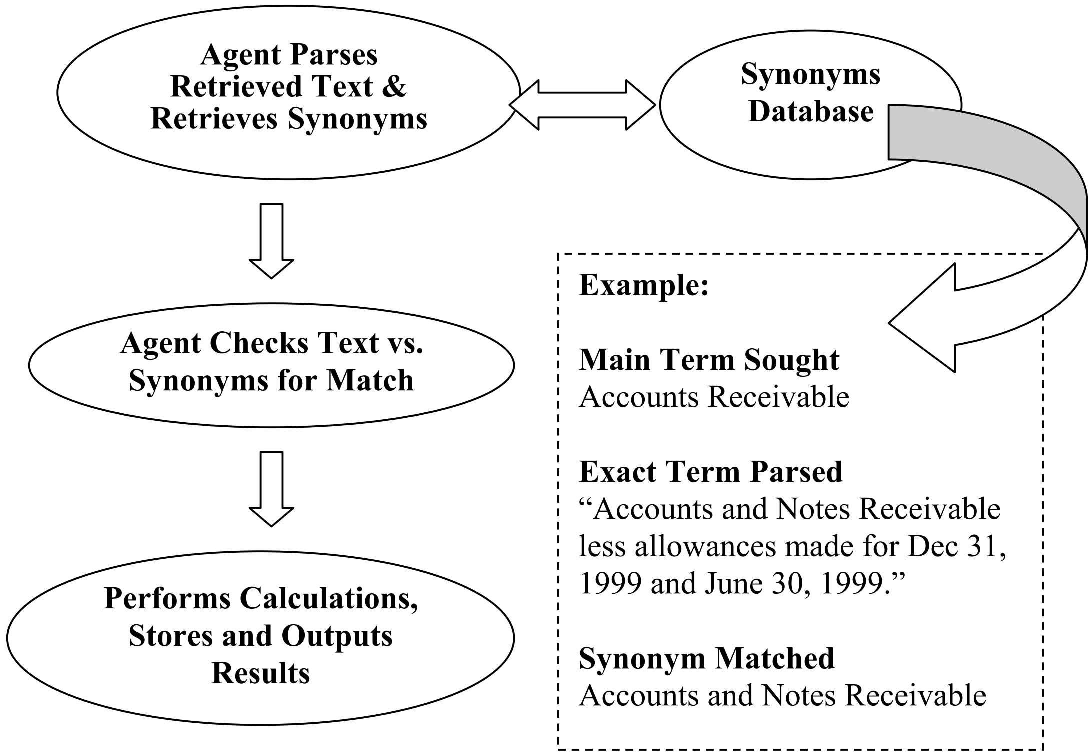
Since the most difficult task of FRAANK is the intelligent analysis of natural text documents (i.e., 10-Q filings), FRAANK needs a very strong pattern matching capability. Therefore, the programming logic of FRAANK is implemented entirely in Perl (Wall et al. [50]), since arguably, Perl provides one of the strongest pattern matching functionalities. By far the largest part of FRAANK's code is devoted to the implementation of intelligent parsing of semi-structured natural text of the SEC filings.

Since FRAANK interacts with both its users and its external information sources over the Internet, the choice of Perl provides the benefits of very strong networking support. When FRAANK communicates with external information sources over the Internet (e.g., SEC, Yahoo, Quicken, etc.), FRAANK essentially acts as a specialized Web client. In the early days of the project, when FRAANK consisted of just the Edgar sub-agent and few mature high level networking modules were available in Perl, FRAANK used a low level approach to implementing the Web client functionality by relying on Perl's implementation of the Berkeley Sockets library to establish HTTP connections with Web sites. This required manual coding of HTTP requests in the direct socket-based implementation of the HTTP interaction, which made it very difficult to implement the growing complexity of interaction of FRAANK with external Web-based information sources. Therefore, the Web client functionality of FRAANK was transitioned to a higher level implementation using the libwww-perl library ([31], used to be called LWP, see also Wong [52]), thus allowing a much simpler implementation of more sophisticated Web client functionality in the current version of FRAANK.

End users also interact with FRAANK using HTTP. In this case, however, the user's Web client contacts FRAANK's Web server, which launches FRAANK using Common Gateway Interface (CGI). FRAANK relies on the CGI.pm module developed by Lincoln D. Stein [44], which uses objects to create Web fill-out forms on the fly and to parse their contents. When executed by the HTTP server, the agent checks whether the user has sent any parameters, which are received as CGI URL-encoded variables using the CGI.pm param() method. If none have been passed, a form is returned to the user's browser prompting for a company name. If the CGI variable specifying the company's name is received, the Edgar sub-agent searches the SEC EDGAR database by making a socket connection to the SEC's Web server. The retrieved list of filings is then returned as output to the user's browser. If the CGI variable 10Q is received, then the Edgar sub-agent retrieves the user-specified 10-Q filing from the SEC EDGAR repository and parses it line by line.

The intelligent parsing procedure in the Edgar sub-agent searches for the expected accounting terms used in the analyses. As illustrated in Figure 2, the Edgar sub-agent utilizes a corresponding source of accounting synonyms, stored in a relational database, to cope with the variation encountered in corporate use of terminology and subsequently identify and parse the appropriate substitute terms. The connection to the database is established over ODBC, and the queries are implemented in standard SQL.

Figure 2: Synonym Matching in FRAANK



Synonyms for a subsection of the 10-Q are retrieved and compared against parsed text until a match is found or all synonyms exhausted. The matching synonym is related to its parent term and the corresponding value parsed and stored such that the parent term, the synonym, and the parsed and original text can be retrieved. Subsequent lines of text are parsed until the entire 10-Q is completed. Text without matching main or synonym terms is noted and stored along with its corresponding value for future evaluation as a possible synonym. Calculations without available values are flagged for the user's attention.

5 FRAANK and XBRL

The development of the eXtensible Business Reporting Language (XBRL) has numerous important implications for the FRAANK project. On the one hand, if XBRL is widely adopted, the most difficult task performed by FRAANK – the extraction of accounting numbers from the financial statements – will become so much simpler that a system with the current FRAANK functionality, while remaining a useful decision aid, will lose its claim of being intelligent. However, in that case (which may take a very long while to occur, if it does) the current capabilities of FRAANK will remain extremely useful for processing archival statements, which will probably not be available in XBRL.

On the other hand, the taxonomies developed for XBRL can be utilized to enhance the intelligence of FRAANK in its primary area of activity – processing semi-structured natural text of financial statements. The parsing procedure of the Edgar sub-agent is now completely redesigned as compared with the initial implementation ([36]) to incorporate the available C&I XBRL Taxonomy capturing the US GAAP.

Since the original Edgar agent used a set of key line items chosen on the basis of their highest empirically observed frequency, the list of such items was necessarily limited, and the agent concentrated on accurately parsing a few most important accounting numbers. The parsing subroutines were specially designed for each of the key items parsed. While this approach proved to be quite accurate (see [36]) when limited to the most important accounting numbers, it turned out difficult to scale up to all the line items in the financial statements.

In the current implementation of the Edgar XBRL sub-agent, a unified parsing subroutine is developed, which parses all the line items in exactly the same way. The 552 elements of the C&I XBRL taxonomy, for the balance sheet, income statement and statement of cash flow (Table 5), are used as the key items in a growing database of (currently) approximately 3200 synonyms. This development is likely to ultimately both improve the accuracy of FRAANK's parsing and to identify possible deficiencies in, or necessary extensions to, the XBRL taxonomy[4]. Indeed, the agent keeps track of the exceptions when the parsing subroutine fails, which can happen for two different principle reasons. The first happens when no synonym in the database matches the line item description. After human review of the exceptions log, this synonym will be added to the database with its paired corresponding key, thus eliminating this instance of failure in the future. The second reason for a parsing failure is the absence of the corresponding key item in the database of synonyms. This indicates a gap in the underlying XBRL taxonomy. The identification of such instances in the exceptions log should provide helpful information for the improvement of the underlying XBRL taxonomy, or for identifying industries or firms where specific extensions to the core XBRL taxonomy may be warranted[5]. For example, financial statements from 80 companies in 12 industries, parsed with FRAANK, show proportions of matched line items by industry that range from 54% to 70% (Table 3), and by-statement matching proportions of approximately 64% (Table 4). Factorial analysis of these results shows that the percentage of items matched for the Airlines industry is significantly different from 7 out of 11 of the remaining industries, suggesting a possible candidate for an industry-specific XBRL-taxonomy extension (Table 1), as well as a possible need for additional synonyms in the knowledge-base. As discussed elsewhere (Bovee et al., 2001 [8]), Table 2 shows significant mean differences in the proportion of line items matched by financial

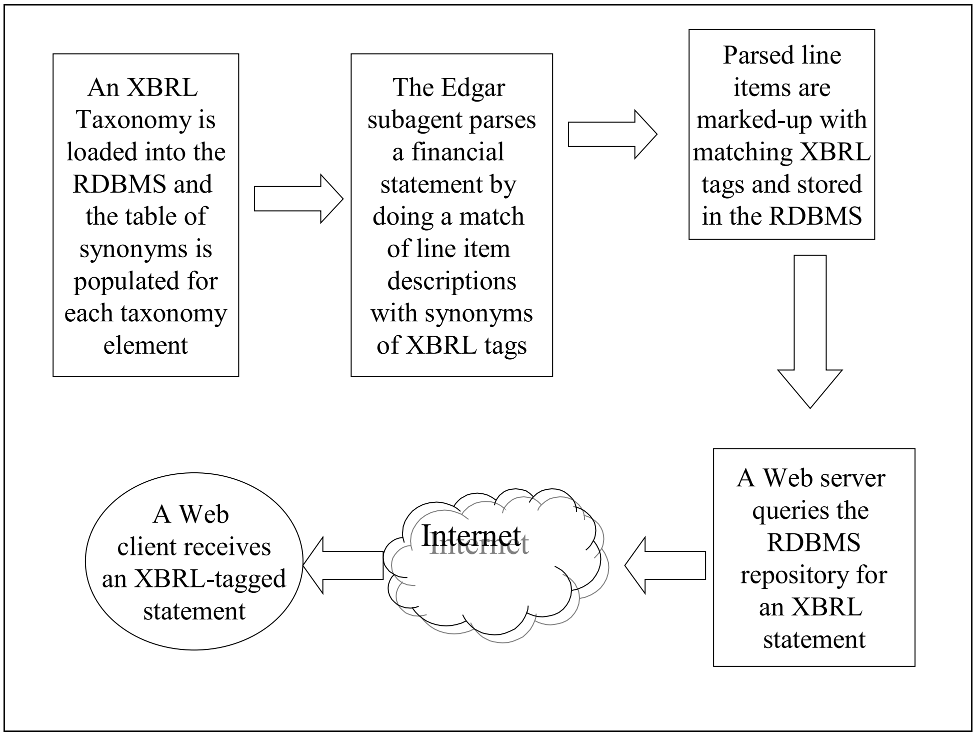
[4] Currently the agent matches approximately 63% of ***all*** line items (see Table 3 and Table 4 for details).

statement, suggesting possible need for examining the XBRL taxonomy, the knowledge-base of synonyms, or both.

The current implementation of the Edgar sub-agent can be used for creating a repository of XBRL-tagged financial statements (e.g. Balance Sheet or Income Statement) or facilitate migration to using XBRL. Since we can presume a high degree of consistency in the format and terminology used for a company between reporting periods, statements tagged by FRAANK can serve as early templates for firms wishing to begin adopting XBRL. While eventually, financial statement tagging will be made at the level of ERP or legacy system, in the early stages of XBRL deployment, the creation of repositories as well as testing of XBRL features will have to be drawn from published financial statements. FRAANK is the enabling technology for this purpose. Figure 3 illustrates an architecture that can be used to extract data from the SEC filings and place them in an XBRL repository. The exception identified by the Edgar sub-agent can be analyzed for commonalities, which can be used to develop industry-specific XBRL taxonomies.

Figure 3: Retrieval of XBRL-tagged statements

[5] The need for industry-specific extensions has already been demonstrated (Bovee et al., 2001 [8]).



6 FRAANK's Applications

In this section we describe the application of FRAANK for various decisions based on financial information.[6] In particular, we outline possible uses of FRAANK that can facilitate making the following decisions online: investment, financial health and creditworthiness evaluation, analytical review procedures in auditing, benchmarking and competitive intelligence.

The help provided by FRAANK is available online and in real-time. Therefore, a decision maker can use FRAANK from any place where the Internet access is available. FRAANK will provide this valuable help right on the user's desktop, and the information will always be as timely as the data sources that are currently publicly available. When online continuous financial reporting becomes available, FRAANK can easily incorporate that source of financial information to provide absolutely timely analysis.

Investment

Internet-based online trading has recently become an important force in the financial markets and investment decision-making. An Internet trader may have to make a fast decision about either taking or liquidating a certain equity position. While watching how the trading in that security develops, this trader can simultaneously use FRAANK to automatically find online, retrieve and integrate relevant information, and compute various ratios as described in the previous section. Savvy investors and financial analysts typically do such analysis manually. Since FRAANK does all this automatically, without explicit instructions from the user, it relieves the user from the relatively routine and time-consuming part of his/her task, and expeditiously allows him/her to concentrate on more intelligent aspects of the decision.

Financial Health and Creditworthiness

A creditor can use FRAANK to facilitate a credit granting decision, especially the quantitative analysis portion. For example, a loan officer will typically want to see how liquid the company is, and whether this company is not in financial distress. It is therefore very helpful for this officer to instantly obtain through FRAANK the net income, the quick and current ratios, the Z-factor for the company, its λ measure of

---

[6] As discussed in Section 7, future developments of FRAANK will incorporate non-financial information

liquidity, and the corresponding risks of bankruptcy and insolvency. The Z-factor (Altman [1, 2]) provides an estimate of bankruptcy risk based on the equation below, derived by multivariate discriminant analysis of bankrupt and non-bankrupt companies.

Altman's Z-Factor

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.010X_5$$

where Z = the overall discriminant score
X_1 = working capital/total assets
X_2 = retained earnings (balance sheet)/ total assets
X_3 = earnings before interest and taxes/total assets
X_4 = market value of equity/book value of total debt
X_5 = sales/total assets

Five ratios and their factors representing measures of corporate liquidity (X_1), solvency (X_2), profitability (X_3, X_4) and activity (X_5) combine to form the Z-factor score. FRAANK must parse each numerator and denominator term in the equation from a line of the appropriate 10-Q section, including the necessary resolution of ambiguous terms with synonyms in the database. A Z-factor score above 2.675/below 1.81 has been shown to indicate low/high risk of bankruptcy (Altman [1, 2]).

The lambda (λ) index (Emery and Cogger [14]) provides a stochastic estimate of the risk of cash and cash reserves being exhausted across a time horizon of interest. As detailed below, the numerator of the lambda index represents cash reserves and expected cash flow for some time horizon at T_0, the beginning of a time horizon of interest. The denominator estimates the size of the typical cash surprise (the square root of the product of the variance in cash flow and the number of time periods in the horizon of interest).

Lambda (λ) Index of Liquidity

use.

$$\lambda = (L_0 + \mu T)/[\sigma(T)^{1/2}]$$

where L_0 = initial liquid reserve at T_0
μ = mean of net cash flow per unit time
σ = standard deviation of net cash flow per unit time
T = length of the period in units of time

FRAANK parses and assembles the cash flow information from 20 sequential quarters for the company[7]. From this set the agent calculates the cash flow mean and standard deviation, and from these derives the λ index, the risk of insolvency, across the time horizon. Intuitively, a company with low initial cash reserves, low mean cash flow per period, or high variability in cash outflow runs a higher risk of becoming insolvent during the period of interest. Additionally, the λ index can be used in a standard normal cumulative distribution table to provide an approximation of the risk of the total of cash reserves and cash flow being exhausted during the time horizon of interest (Emery and Cogger [14], Cogger [15]). For example, a company with total cash for time horizon period T that is twice the typical cash surprise for that period has an approximate probability of cash insolvency during T of:

$$P(\text{insolvency during T}) = 1 - N(\lambda) = 1 - 0.97725 = 0.02275,$$

where $N(\lambda)$ is the cumulative standard normal distribution value associated with λ [15].

Analytical Review Procedures

Auditors are required to perform various analytical procedures both in the planning and completion phases of financial audit. In the planning phase, the purpose of

[7] Where companies do not have 20 quarters of cash flow information, or the 20 are not sequential, FRAANK reports this as an exception. A future modification will permit fewer than 20 quarters at the

the analytical procedures is "... to assist the auditor in planning the nature, timing, and extent of other auditing procedures" (AICPA 1998, AU 329.05 [5]). Such an analysis helps auditors understand client's business and identify areas that may represent specific risks relevant to the audit. Analytical procedures involve identifying unusual trends in the account balances and their relationships by performing ratio analysis such as quick and current ratios, inventory turnover, gross margin ratio, etc., and comparing these ratios with both industry standards, and with those of peers and competitors, as well as with past performance.

FRAANK can help the auditor by automatically gathering online the relevant financial information about the client's peers, calculating the appropriate ratios, and presenting them to the auditor. This frees the auditor from the mundane tasks, and enables him/her to concentrate on more important and difficult issues.

Benchmarking and Competitive Intelligence

Management also compares financial ratios with industry averages and with those of peers, suppliers and competitors to monitor and evaluate company performance and identify competitive threats and advantages. As described above, FRAANK can again be used as an invaluable online tool for this intelligence-gathering purpose. Using FRAANK, companies can quickly compare themselves to competitors in their industry sector or market across a variety of standard financial ratios. For example, a company that finds its age of accounts receivable is 60 days, while that of its industry competitors is 30 days, may need to change its payment policy to enhance cash flow, or establish a

user's discretion.

carrying charge premium for its more generous payment terms. Conversely, if its age of accounts receivable is 30 days while that of its competitors is 60 days, it may be losing business due to more stringent payment terms. If its age of accounts receivable is the same as competitors', it may be able to *offer* better terms at a premium, again to enhance cash flow and attract more business.

Similarly, the λ index permits both benchmarking against other companies in a given industry and pinpointing potential competitor's cash flow problems. A competitor's low λ index may signal their higher risk of sudden costs or investments exceeding their typical cash outflow, an impending need for outside capital support, or an overall high risk of insolvency. Recent interest in the burn-rate (cash reserves versus operating expenses) of Internet companies [51] has demonstrated that, in the accelerated world of so-called dot.com companies, both individual investors and professional analysts exhibit intense interest in such estimates. However, since the λ index also incorporates variability in cash flow, it provides additional information on the risk of insolvency beyond that of a linear estimate of cash expenditures versus cash reserves. FRAANK can be integrated with a management information system to automate this monitoring and evaluation process by retrieving the necessary accounting numbers from public Internet repositories and doing the necessary calculations.

7 Future Developments

The FRAANK agent lays down the foundation for additional research in the area of online financial reporting and virtual auditing. Direct research resulting from this study will extend and refine the FRAANK agent's capabilities for gathering, isolating and analyzing key financial data from the Edgar database and other information sources. We discuss below several important directions along which FRAANK's capabilities can be significantly enhanced.

Automatic Identification of Synonyms

The current version of FRAANK analyzes line items of SEC filings using a separate program code and a table of synonyms for each line item analyzed. Although this approach allows FRAANK to achieve a high level of accuracy in identifying line items, it is very difficult to make this process 100% accurate without a tremendous amount of human involvement in the expansion of the synonyms database. We therefore plan to research the possibility of using machine learning techniques for automated acquisition of new synonyms of accounting terms found in the text of financial statements, using the structure of the statements. This development should result in automatically expanding the table of synonyms of accounting terms whenever FRAANK encounters a new synonym.

News Portals

Users of FRAANK will be interested in obtaining other types of relevant information not currently available from FRAANK. This includes intelligent digests of the general economic and company specific news. Such information can be found on various information portals like Yahoo or Quicken. Future research and developments of FRAANK in this direction will add news sub-agents and intelligent digesting mechanisms to achieve this goal.

Integration of FRAANK with Stand-alone AI Systems

The intelligence capabilities of FRAANK can be enhanced by integrating it with some of the existing AI systems for accounting and auditing such as neural networks, rule-based systems, or belief networks. For example, one can integrate FRAANK with artificial neural networks to develop models for predicting bankruptcies (Fanning and Cogger 1994 [19]), recognizing potentials for management fraud (Fanning et al. 1995 [20]), and making going concern judgments (Biggs et al. [6]). This integration will result in providing automated timely expert advice about the client and the industry.

Companies' Web Sites

Although most companies' Web sites currently limit presented financial information to what is already available from the Edgar filings, they still contain useful additional information such as press releases. Moreover, future developments in online business reporting will make companies' Web sites even more essential source of relevant information for FRAANK. However, obtaining that information presents a major challenge for the current FRAANK technology, since the structure, content and format of corporate Web sites all vary greatly, are subject to change without notice, and the number of companies is very large. It will therefore be extremely important to research the possibility of developing new technology for finding important financial and non-financial information in companies' Web sites of arbitrary structure. This technology will probably utilize formal representation and learning of the structure of Web information sources.

8 Concluding Remarks

This paper has described the design and implementation of FRAANK – the intelligent agent that gathers financial information about publicly traded companies over the Internet and then processes it to help various decision makers (e.g., investors, creditors, auditors, managers). More specifically, FRAANK finds and understands important accounting numbers in the quarterly filings available from the SEC EDGAR online repository, and integrates these numbers with the current stock quotes and consensus earnings per share forecasts to calculate various financial ratios and financial health indicators (such as Altman's Z-factor and the λ index).

The Internet and electronic commerce will undoubtedly bring about more frequent reporting of financial information, thus creating the need for the more frequent auditing of these financial statements. FRAANK and similar tools provide a test bed for evaluating processes and metrics needed to accomplish these tasks and will make them feasible by drastically reducing both the cost and the duration of the audit.

References

1. Altman, Edward I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, Vol. 23, no. 4, September 1968.

2. Altman, Edward I. 1983. *Corporate Financial Distress: A complete guide to predicting, avoiding and dealing with bankruptcy.* John Wiley & Sons, New York, NY.

3. American Institute of Certified Public Accountants (AICPA SCAS). Report of the Special Committee on Assurance Services:

http://www.aicpa.org/assurance/scas/index.htm

4. American Institute of Certified Public Accountants (AICPA WebTrust). CPA WebTrust: http://www.aicpa.org/webtrust/index.htm

5. American Institute of Certified Public Accountants (AICPA, 1998) AICPA Professional Standards, Vol. 1, AU Section 329, 1998.

6. Biggs, S. F., M. Selfridge and G. R. Krupka. 1992. A computational model of auditor knowledge and reasoning process in the going-concern judgment. *Auditing: A Journal of Practice and Theory, Supplement*, pp. 82-112.

7. Boritz, E., E. Mackler and D. McPhie. 1999. Reporting on systems reliability. *Journal of Accountancy*, Vol. 188, No. 5, November. http://www.aicpa.org/pubs/jofa/nov1999/boritz.html

8. Bovee M., M. Ettredge, R.P. Srivastava, M.A. Vasarhelyi. 2001. *Enabling Semantic Consistency for Electronic Business Reporting: Does the Year 2000 XBRL Taxonomy Accommodate Current Business Financial Reporting Practice?* Working paper.

9. Brown, C., L. Gasser, D. O'Leary, and A. Sangster. 1995. AI on the WWW: Supply and demand agents. *IEEE Expert*, Vol. 10, No. 4, August, pp. 50-55.

10. Doorenbos, R.B., O. Etzioni, and D. Weld. 1997. A scalable comparison-shopping agent for the World-Wide Web. *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, Association for Computing Machinery, New York, pp. 39-48: ftp://ftp.cs.washington.edu/pub/etzioni/softbots/agents97.ps

11. EDGAR Database, U.S. Securities and Exchange Commission: http://www.sec.gov/edgarhp.htm

12. EdgarScan, PwC Global Technology Centre: http://edgarscan.tc.pw.com/

13. Elliott, Robert K. 1997. Assurance service opportunities: Implications for academia. *Accounting Horizons*, Vol. 11, No. 4, December, pp. 61-74.

14. Emery, G. W. and K. O. Cogger. 1982. The Measurement of Liquidity. *Journal of Accounting Research*, Autumn, Pt. 1, pp. 290-302.

15. Emery, G. W. 1999. *Measuring Credit Risk Using Lambda Analysis*. Working Paper, February.

16. Etzioni, O. and D. Weld. 1994. A softbot-based interface to the Internet. *Communications of the ACM*, Vol. 37, No. 7, July, pp. 72-76: ftp://ftp.cs.washington.edu/pub/etzioni/softbots/cacm.ps

17. Etzioni, O. and D. Weld. 1995. Intelligent agents on the Internet: Fact, fiction and forecast. *IEEE Expert*, Vol. 10, No. 4, August, pp. 44-49:

ftp://ftp.cs.washington.edu/pub/ai/ieee-expert.ps.Z

18. Etzioni, O. 1997. Moving up the information food chain: Deploying softbots on the World Wide Web. *AI Magazine*, Vol. 18, No. 2, Summer, pp. 11-18: ftp://ftp.cs.washington.edu/pub/etzioni/softbots/a96.ps.gz

19. Fanning, K. and K. Cogger. 1994. A comparison analysis of artificial neural networks for financial distress. *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 3, No. 4, December, pp. 241-252.

20. Fanning, K., K. Cogger and R. Srivastava. 1995. Detection of management fraud: A neural network approach. *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 4, pp. 113-126.

21. Ferguson, Don. 1997. Parsing financial statements efficiently and accurately using C and Prolog. *The Fifth International Conference and Exhibition on the Practical Applications of Prolog*, April 1997, London, UK: http://www.practical-applications.co.uk/PAP97/

22. Foundation for Intelligent Physical Agents (FIPA): http://www.fipa.org/

23. Franklin, Stan and Art Graesser. 1997. Is it an agent, or just a program?: A taxonomy for autonomous agents. In: Intelligent Agents III. Agent Theories, Architectures, and Languages. Eds.J.P. Müller, M.J. Wooldridge, N.R. Jennings. *Lecture Notes in Computer Science*, Vol. 1193. Springer: http://www.msci.memphis.edu/~franklin/AgentProg.html

24. Genesereth, M. and S. Ketchpel. 1994. Software agents. *Communications of the ACM*, Vol. 37, No. 7, July, pp. 48-53: http://logic.stanford.edu/sharing/papers/agents.ps

25. Greenstein, Marilyn. 1998. Web site attestation – A study of licensed providers. *Collected Papers of the Seventh Annual Research Workshop on: Artificial Intelligence and Emerging Technologies in Accounting, Auditing and Tax*. Edited by Carol E. Brown, American Accounting Association, New Orleans, LA, August 1998, 43-56.

26. Huhns, M.N. and M.P. Singh. 1997. Agents on the Web: Conversational agents. *IEEE Internet Computing*, Vol. 1, No. 2, pp. 73-75.

27. Huhns, M.N. and M.P. Singh. 1997. Agents on the Web: Mobile agents. *IEEE Internet Computing*, Vol. 1, No. 3, pp. 80-82.

28. Jain, A.K., M. Aparicio IV and M.P. Singh. 1999. Agents for process coherence in virtual enterprises. *Communications of the ACM*, Vol. 42, No. 3, March, pp. 62-69.

29. Lange, D.B. and M. Oshima. 1999. Seven good reasons for mobile agents. *Communications of the ACM*, Vol. 42, No. 3, March, pp. 88-89.

30. Lea, Richard. 1997. Opportunities for assurance services in the 21st century: A progress report of the Special Committee on Assurance Services. *Proceedings of the 1996 Deloitte and Touche/University of Kansas Symposium on Auditing Problems*, May, pp. 1-36.

31. libwww-perl Library: http://www.linpro.no/lwp/

32. Kotz, David and Robert S. Gray. 1999. Mobile Agents and the Future of the Internet. *ACM Operating Systems Review*, **33** (3), August, pp. 7-13.

33. Maes, Pattie. 1994. Agents that reduce work and information overload. *Communications of the ACM*, Vol. 37, No. 7, July, pp. 30-40.

34. Maes, Pattie, R.H. Guttman, and A.G. Moukas. 1999. Agents that buy and sell. *Communications of the ACM*, Vol. 42, No. 3, March, pp. 81-87,90-91.

35. Mui, C. and W.E. McCarthy. 1987. FSA: Applying AI techniques to the familiarization phase of financial decision making. *IEEE Expert*, Vol. 2, No. 3, Fall, pp. 33-41.

36. Nelson, Kay M., A. Kogan, R.P. Srivastava, M.A. Vasarhelyi, and H. Lu. 2000. Virtual auditing agents: the EDGAR Agent challenge. *Decision Support Systems*, **28** (3), pp. 241-253.

37. O'Leary, Daniel E. 1997. Using AI in accounting systems to facilitate electronic commerce: Intelligent agents and virtual organizations. *Rutgers Colloquium on Accounting 2000: Electronic Commerce, Artificial Intelligence and Other Information Technologies*, May, Newark, NJ.

38. O'Leary, Daniel E. and T. Munakata. 1988. Developing consolidated financial statements using a prototype expert system. In: *Applied Expert Systems*, E. Turban and P.R. Watkins (Editors), pp. 143-157. Elsevier Science Publishers, Amsterdam.

39. O'Leary, Daniel E. and T. Eis. 1991. A knowledge-based system for cash management: With implications for structuring DSS and with extensions for system learning and knowledge acquisition. *Advances in Working Capital Management*, JAI Press, Vol. 2, pp. 197-209.

40. O'Leary, Daniel E. and N. Kandelin. 1992. ACCOUNTANT: A domain dependent accounting language processing system. In: *Expert Systems in Finance*, D.E. O'Leary and P.R. Watkins (Editors), pp. 253-267. Elsevier Science Publishers, Amsterdam.

41. Petrie, Charles J. 1996. Agent-based engineering, the Web, and intelligence. *IEEE Expert*, Vol. 11, No. 6, December, pp. 24-29: http://cdr.stanford.edu/NextLink/Expert.html

42. Quicken.com Investments: http://www.quicken.com/investments

43. Singh, M.P. 1998. Agent communication languages: Rethinking the principles. *IEEE Computer*, Vol. 31, No. 12, December, pp. 40-47.

44. Stein, Lincoln D. 1999. CGI.pm - a Perl5 CGI Library: http://stein.cshl.org/WWW/software/CGI/cgi_docs.html

45. Steier, David. 1995. Comparable datasets in performance benchmarking. In: *Working Notes of the 1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, ed. Craig Knoblock & Alon Levy. Also available as Technical Report 34, PricewaterhouseCoopers Technology Centre: http://www.pwcglobal.com/extweb/ncovrvw.nsf/0cc1191c627d157d8525650600609c03/0b087e9b819544cb852566ad005153d6/$FILE/Tr34.pdf

46. Steier, David M., Scott B. Huffman, and Douglas I. Kalish. 1997. Beyond full-text search: AI-based technology to support the knowledge cycle. *AAAI Spring Symposium on AI in Knowledge Management*, March, pp. 161-167. Also available as Technical Report 48, PricewaterhouseCoopers Technology Centre: http://www.pwcglobal.com/extweb/ncovrvw.nsf/0cc1191c627d157d8525650600609c03/0b087e9b819544cb852566ad005153d6/$FILE/tr48.pdf

47. Tanaka, S. 1982. *The Structure of Accounting Language.* Chuo University Press Tokyo, Japan.

48. UMBC KQML Web: http://www.cs.umbc.edu/kqml/

49. Vasarhelyi, Miklos A. and F.B. Halper. 1991. The continuous audit of online systems, *Auditing: A Journal of Practice and Theory*, Vol. 10, No. 1, Spring, pp. 110-125.

50. Wall, Larry, Tom Christiansen and Randal L. Schwartz. 1996. *Programming Perl*, 2nd Edition, O'Reilly and Associates, Sebastopol, CA.

51. Willoughby, Jack. 2000. Burning Up. *Barron's*, March 20, 2000, pp. 29-32.

52. Wong, Clinton. 1997. *Web Client Programming with Perl: Automating Tasks on the Web*, O'Reilly and Associates, Sebastopol, CA.

53. Wong, David, N. Paciorek, and D. Moore. 1999. Java-based mobile agents. *Communications of the ACM*, Vol. 42, No. 3, March, pp. 92-102.

54. Wooldridge, Michael. 1999. Intelligent agents. In *Multiagent Systems*, ed. G. Weiss, The MIT Press, April http://www.csc.liv.ac.uk/~mjw/pubs/mas99.ps.gz

55. Wooldridge, Michael and Nicholas R. Jennings. 1995. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, Vol. 10, No. 2, pp. 115-152: http://www.doc.mmu.ac.uk/STAFF/mike/ker95.ps

56. Yahoo! Finance: http://quote.yahoo.com/

Appendix

Table 1 Mean differences in proportion by industry. Cells show the difference (I-J) for proportion of matched financial statement terms, with significant differences bolded. For example, mean proportion of matched line times for Beverages minus mean proportion matched for Airlines = 9.379 percent, and is significant. Airlines significantly differs from many industries, suggesting a candidate for an industry-specific XBRL taxonomy extension.

		J										
		Airlines	Automotive Retail & Services	Beverages	Computers & Office Equipt.	Computer Software	Entertainment	Food & Drug Stores	Food Services	General Merchandisers	Motor Vehicles & Parts	Petroleum
I	Automotive Retail & Services	8.051	-									
	Beverages	**9.379**	1.329	-								
	Computers & Office Equipt.	6.393	(1.658)	(2.987)	-							
	Computer Software	**14.494**	6.443	5.114	8.101	-						
	Entertainment	3.161	(4.890)	(6.219)	(3.232)	**(11.333)**	-					
	Food & Drug Stores	6.645	(1.405)	(2.734)	0.253	(7.849)	3.485	-				
	Food Services	**15.848**	7.797	6.468	**9.455**	1.354	**12.687**	**9.203**	-			
	General Merchandisers	**8.446**	0.395	(0.933)	2.053	(6.048)	5.285	1.801	(7.402)	-		
	Motor Vehicles & Parts	**16.631**	8.581	7.252	**10.239**	2.137	**13.471**	**9.986**	0.783	8.185	-	
	Petroleum	**10.808**	2.758	1.429	4.416	(3.686)	7.648	4.163	(5.040)	2.362	(5.823)	-
	Pharmaceuticals	**11.918**	3.868	2.539	5.526	(2.576)	**8.758**	5.273	(3.930)	3.472	(4.713)	1.110

Table 2 Mean differences in proportion matched by statement. Cells show the difference (I-J) for proportion of matched financial statement terms, with significant differences bolded. For example, mean proportion of matched line times for Balance Sheet minus mean proportion matched for Statement of Cash Flow = 4.422 percent, and is significant.

		J	
		Income Statement	Statement of Cash Flow
I	Balance Sheet	4.139	**4.422**
	Income Statement		0.283

Table 3 Means and standard errors for proportions of line items matched per industry.

	Mean	Std. Error	95% Confidence Interval	
INDUSTRY			Lower Bound	Upper Bound
Airlines	53.945	2.909	48.209	59.681
Automotive Retail & Services	61.995	2.909	56.259	67.731
Beverages	63.324	2.909	57.588	69.060
Computers & Office Equipt	60.337	2.909	54.601	66.073
Computer Software	68.438	2.909	62.702	74.174
Entertainment	57.105	2.909	51.369	62.841
Food & Drug Stores	60.590	2.909	54.854	66.326
Food Services	69.792	3.142	63.597	75.988
General Merchandisers	62.391	2.909	56.655	68.127
Motor Vehicles & Parts	70.576	3.849	62.988	78.164
Petroleum Refining	64.753	2.909	59.017	70.489
Pharmaceuticals	65.863	2.909	60.127	71.599

Table 4 Mean and standard errors for proportion of line items matched by statement type.

	Mean	Std. Error	95% Confidence Interval	
Statement			Lower Bound	Upper Bound
BS	66.113	1.509	63.137	69.089
IS	61.974	1.509	58.998	64.949
CFS	61.691	1.509	58.715	64.666

Grand Mean Proportion of line items matched.

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
63.259	.871	61.541	64.977

Table 5 Numbers of XBRL tags and Knowledge-base synonyms by XBRL taxonomy level.

Root level is the top tag in the taxonomy, i.e. "statement".

BALANCE SHEET	LEVEL	# TAGS	# SYNONYMS
	1	1	1
	2	1	4
	3	2	6
	4	9	42
	5	52	134
	6	76	159
	7	100	145
	8	30	44
	9	4	4
TOTAL COUNT		**275**	**539**

INCOME STATEMENT	LEVEL	# TAGS	# SYNONYMS
	1	1	1
	2	1	1
	3	2	24
	4	6	19
	5	9	108
	6	17	170
	7	18	40
	8	5	52
	9	4	19
	10	18	101
	11	14	123
	12	19	121
	13	17	79
TOTAL		**131**	**858**

CASH FLOW STATEMENT	LEVEL	# TAGS	# SYNONYMS
	1	1	1
	2	1	1
	3	2	35
	4	5	104
	5	7	124
	6	10	114
	7	42	618
	8	43	489
	9	25	267
	10	10	10
TOTAL		**146**	**1763**